ESTIMATING PRESENCE OR ABSENCE OF SMOKING THROUGH BIOSIGNALS

1. INTRODUCTION

1.1 OVERVIEW

Smoking is a major public health concern worldwide, with more than 1 billion people estimated to be regular smokers. It is a major contributor to avoidable deaths and is linked to several health issues, such as lung cancer, heart disease, stroke, and respiratory illnesses. Therefore, understanding the physical and psychological effects of smoking on the body is important for developing effective interventions to help people quit smoking.

Researchers have conducted studies on the consequences of smoking on the human body to better understand the physical signals connected with smoking. The signals that have been examined in these investigations include, among others, respiratory function, cardiovascular health, and cognitive performance.

Our analysis will focus on understanding how information of different body signals can be used to determine if a person is a smoker or not. The dataset used for this project has been taken from Kaggle and is titled by smoking. The motivation for choosing this topic stems from a keen interest in public health and a desire to uncover potential insights from the data. By analyzing the body signals of smokers, we aim to gain a deeper understanding of the health risks associated with smoking and contribute to the development of effective smoking cessation strategies.

1.2 PURPOSE

The main aim or goal of a machine learning project estimating the presence or absence of smoking through bio-signals is to develop an accurate and reliable system that can detect smoking behavior using physiological signals.

The system would use machine learning algorithms to analyze bio-signals, such as heart rate, skin conductance, and respiratory rate, to determine if a person is smoking or not. The system could be applied in various contexts, such as clinical settings or workplaces, to monitor smoking behavior and help individuals quit smoking.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM

There are several existing methods to identify the presence or absence of smoking in the body, although they may not always be efficient or accurate. Some of these methods include:

Self-reporting: Individuals can simply admit to smoking when asked about their smoking habits. However, this method relies on honesty and may not always be reliable, as some people may underreport or deny their smoking habits.

Breath tests: Breathalyzer tests can detect the presence of carbon monoxide (CO) or other chemicals associated with smoking in a person's breath. Elevated CO levels are indicative of recent smoking. However, this method only provides information about recent smoking and not long-term habits.

Urine tests: Nicotine and its metabolites can be detected in urine samples. These tests can reveal recent nicotine exposure, but like breath tests, they do not provide insights into long-term smoking habits.

Blood tests: Blood tests can also measure nicotine and its metabolites, offering insights into recent smoking. However, they are not well-suited for long-term monitoring.

Biomarker analysis: Some biomarkers in the body, such as cotinine (a nicotine metabolite) or certain proteins, can indicate recent or long-term smoking. These biomarkers can be measured in bodily fluids like blood or saliva.

2.2 PROPOSED SYSTEM

Estimating Presence or Absence of Smoking through Bio Signals" aim to improve the efficiency and accuracy of identifying smoking habits. Here's why such projects can be more efficient:

Continuous Monitoring: Machine learning models can analyze bio signals over time, providing insights into long-term smoking habits, whereas traditional methods often focus on recent exposure.

Pattern Recognition: ML models can recognize patterns in bio signals that are indicative of smoking, even if individuals deny their habits. This reduces reliance on self-reporting.

Data Integration: ML models can integrate multiple types of bio signals (e.g., heart rate, skin conductance, respiratory patterns) to create a more comprehensive and accurate assessment.

Real-time Detection: Some ML applications can detect smoking behaviors in real-time, which can be valuable for immediate interventions or support.

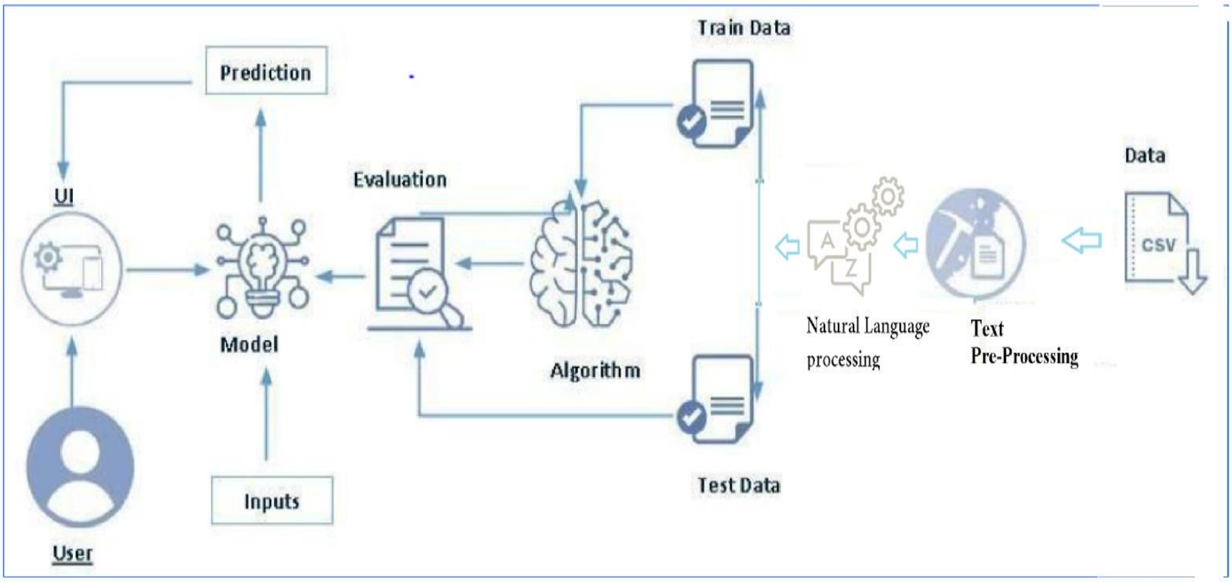
Personalization: ML models can be trained to consider individual variations in bio signals, improving the accuracy of smoking detection for different people.

Privacy Preservation: ML models can be designed to protect user privacy by analyzing bio signals without collecting sensitive information, addressing concerns about data privacy.

In summary, machine learning projects focused on estimating smoking presence or absence through bio signals have the potential to offer more efficient and accurate methods for identifying smoking habits, especially in situations where traditional methods may fall short. These projects leverage advanced algorithms to process and interpret complex data patterns, leading to improved detection and monitoring of smoking behaviors.

3. THEORETICAL ANALYSIS

3.1 BLOCK DIAGRAM



3.2 HARDWARE AND SOFTWARE DESIGNING

Python

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. It was created by Guido van Rossum , and first released on February 20, 1991. Its high-level built in data structures, combined with dynamic typing and dynamic binding , make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readabilityand therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

Anaconda Navigator

Anaconda Navigator is a free and open-source distribution of the Python and R programming languages for data science and machine learning related applications. It can be installed on Windows, Linux, and macOS.Conda is an open-source, cross platform, package management system. Anaconda comes with so very nice tools like JupyterLab, Jupyter Notebook, QtConsole, Spyder, Glueviz, Orange, Rstudio, Visual Studio Code. For this project, we will be using Jupyter notebook and Spyder.

Jupyter Notebook

The Jupyter Notebook is an open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text. Jupyter Notebook is maintained by the people at Project Jupyter. Jupyter Notebooks are a spinoff project from the IPython project, which used to have an IPythonNotebook project itself. The name, Jupyter, comes from the core supported programming languages that it supports: Julia, Python, and R. Jupyter ships with the IPython kernel, which allows you to write your programs in Python, but there are currently over 100 other kernels that you can also use.

Spyder

Spyder, the Scientific Python Development Environment, is a free integrated development environment (IDE) that is included with Anaconda. It includes editing, interactive testing, debugging, and introspection features. Initially created and developed by Pierre Raybaut in 2009, since 2012 Spyder has been maintained and continuously improved by a team of scientific Python developers and the community. Spyder is extensible with first-party and third party plugins includes support for interactive tools for data inspection and embeds Pythonspecific code. Spyder is also pre-installed in Anaconda Navigator, which is included in Anaconda.

Flask

Webframework used for building. It is a web application framework written in python which will be running in local browser with a user interface. In this application, whenever the user interacts with UI and selects emoji, it will suggest the best and top movies of that genre to the user.

Hardware Requirements

Operating system: window 11 with 64 bit

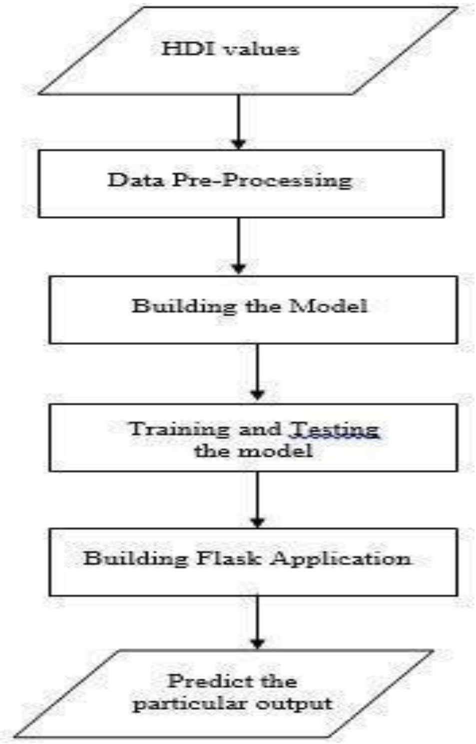
Processor Type -Intel Core i5

RAM: 8GB and above
Hard disk: min 100GB

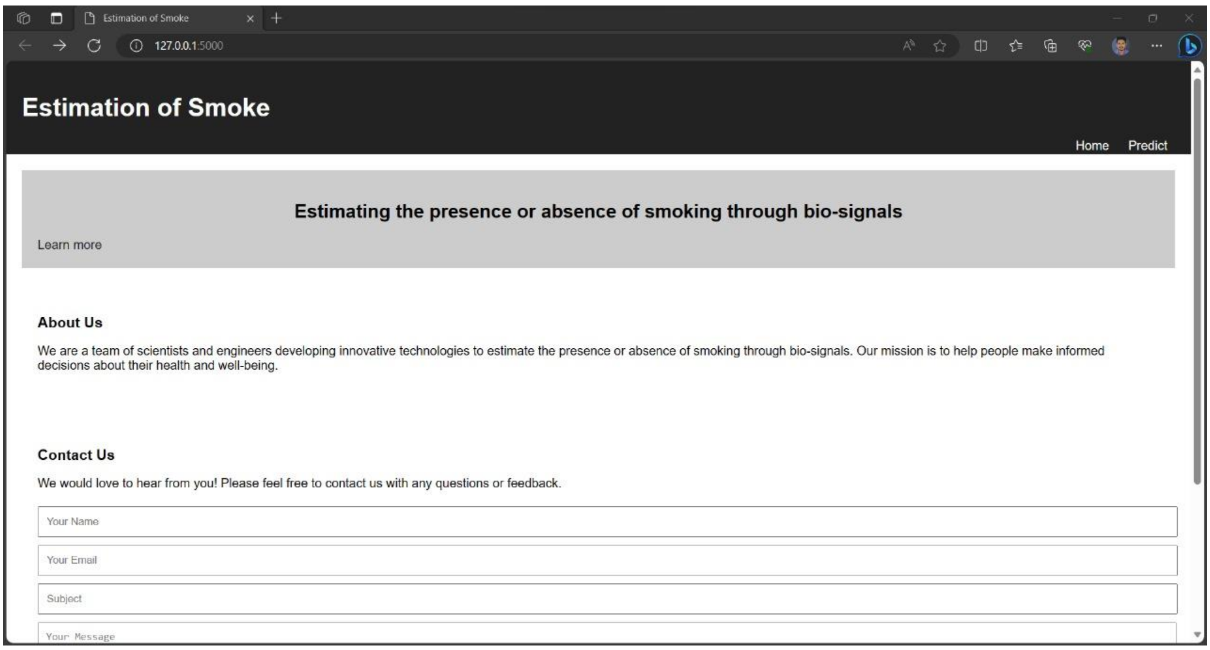
4. EXPERIMENTAL INVESTIGATION

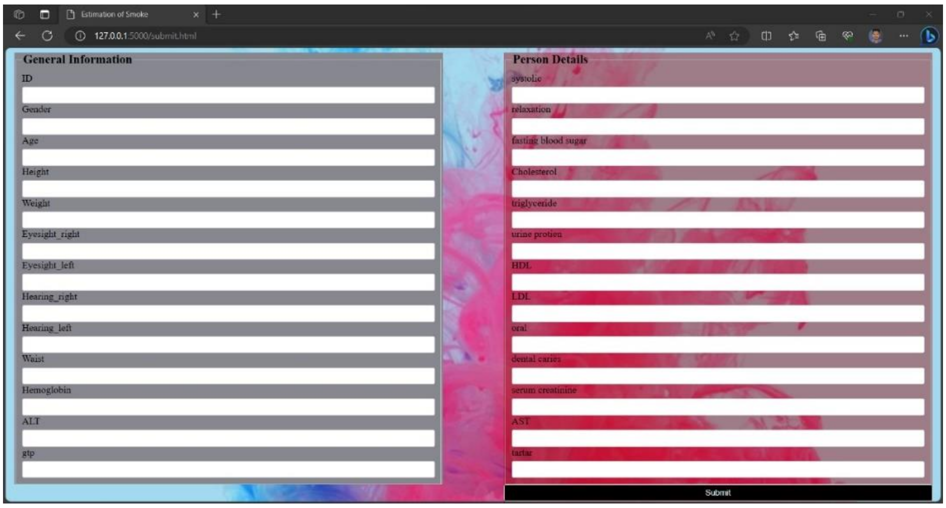
The text data need to be organized before proceeding with the project. The original dataset has a single folder. We will be using the city_day.csv file to fetch the text data of training data. The datas need to be unique and all fields need to be filled. The dataset images are to be pre-processed before giving to the model. We will create a function that uses the pre-trained model for predicting custom outputs. Then we have to test and train the model. After the model is build, we will be integrating it to a web application.

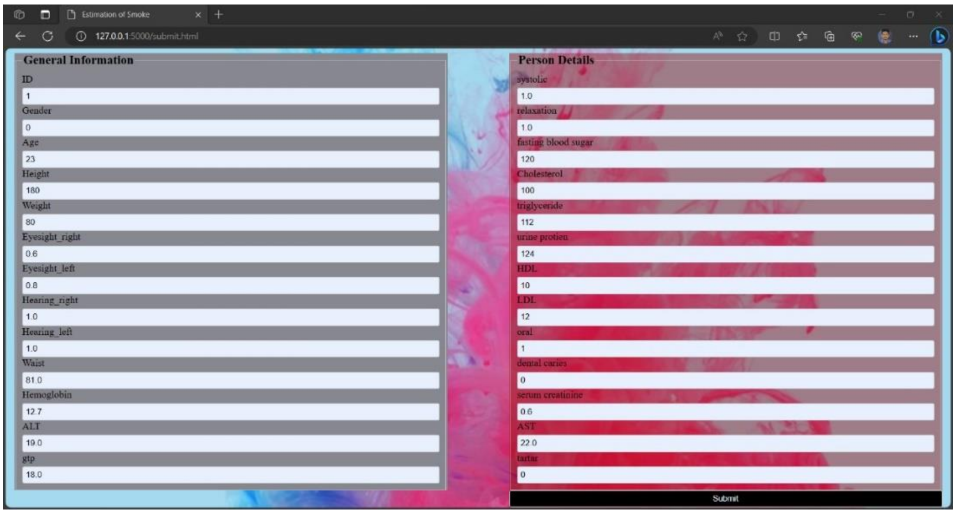
5. FLOWCHART

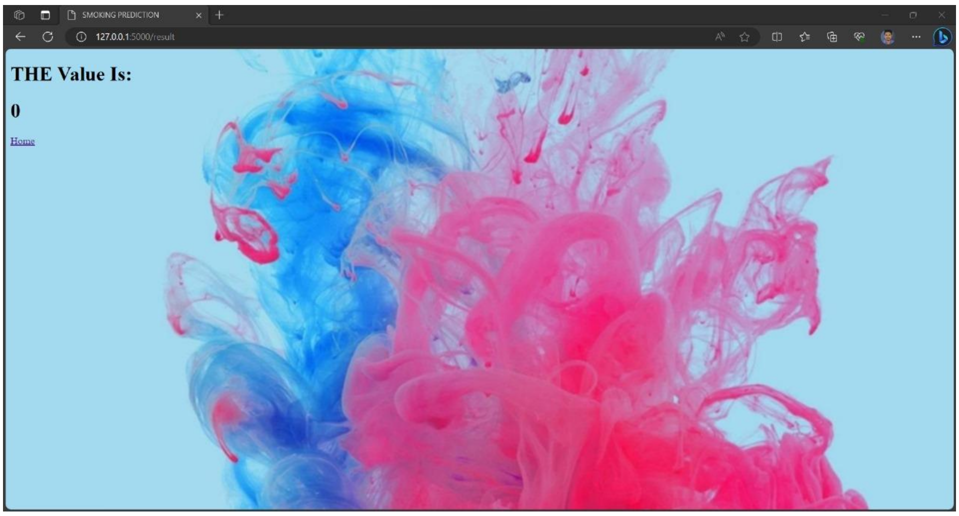


6. RESULT




Estimation of Smoke
127.0.0.1:5000/submit.html
General Information
ID
Gender
Age
Height
Weight
Eyesight_right
Eyesight_left
Hearing_right
Hearing_left
Waist
Hemoglobin
ALT
gtp
Person Details
systolic
relaxation
fasting blood sugar
Cholesterol
triglyceride
urine protien
HDL
LDL
oral
dental caries
serum creatinine
AST
tartar
Submit


Estimation of Smoke
127.0.0.1:5000/submit.html
General Information
ID
1
Gender
0
Age
23
Height
180
Weight
80
Eyesight_right
0.6
Eyesight_left
0.8
Hearing_right
1.0
Hearing_left
1.0
Waist
81.0
Hemoglobin
12.7
ALT
19.0
gtp
18.0
Person Details
systolic
1.0
relaxation
1.0
fasting blood sugar
120
Cholesterol
100
triglyceride
112
urine protien
124
HDL
10
LDL
12
oral
1
dental caries
0
serum creatinine
0.6
AST
22.0
tartar
0
Submit


SMOKING PREDICTION
127.0.0.1:5000/result
THE Value Is:
0
Home

7. ADVANTAGES & DISADVANTAGES

Advantages

Increased Accuracy: Can analyze complex patterns in bio signals, potentially leading to more accurate detection of smoking behavior compared to traditional methods.

Continuous Monitoring: This models can provide ongoing monitoring, allowing for a better understanding of smoking habits over time, which can be essential for intervention and support.

Objective Assessment: Assessments are less reliant on self-reporting, reducing the risk of biased or inaccurate information.

Data Integration: Can integrate multiple types of bio signals (e.g., heart rate, skin conductance, respiratory patterns), providing a more comprehensive assessment of smoking behavior.

Personalization: Can adapt to individual variations in bio signals, improving accuracy for different people.

Disadvantages

Ethical Concerns: Handling sensitive bio data and monitoring individuals' behavior can raise ethical concerns related to privacy and consent.

Costly Implementation: Developing and deploying models for real-time monitoring can be expensive

User Acceptance: Some individuals may be uncomfortable with the idea of constant monitoring of their bio signals, which could affect the adoption and success of the technology.

Limited Generalization: ML models developed in one population or setting may not generalize well to other populations or contexts, limiting their broader applicability.

8. APPLICATIONS

The project "Estimating Presence or Absence of Smoking through Bio Signals" has versatile applications across various domains. Here are several areas where this technology can be applied:

Healthcare and Clinical Settings:

Smoking Cessation Programs: Integrated into smoking cessation programs to monitor and support individuals trying to quit smoking.

Chronic Disease Management: Used to track and manage health conditions exacerbated by smoking, such as cardiovascular diseases and respiratory illnesses.

Public Health Initiatives:

Population Health Monitoring: Applied to assess smoking prevalence and trends in specific regions, aiding public health agencies in targeted interventions.

Anti-Smoking Campaigns: Used to evaluate the effectiveness of anti-smoking campaigns and policies.

Education and Awareness:

Schools and Universities: Integrated into educational programs to raise awareness about the dangers of smoking among students.

Health Education: Included in health education curricula to educate individuals about the impact of smoking on health.

Drug Rehabilitation Centers:

Substance Abuse Treatment: Applied to monitor and support individuals undergoing substance abuse treatment, including smoking cessation.

Prison Systems:

Incarceration Facilities: Used to monitor and discourage smoking among inmates, as smoking-related health issues can be particularly problematic in correctional settings.

Sports and Fitness Industry:

Sports Training: Employed in sports science to assess athletes' smoking habits and their impact on performance.

Fitness Apps: Integrated into fitness apps to discourage smoking among users pursuing healthier lifestyles.

Government Regulations and Compliance:

Tobacco Control: Used for compliance checks in areas where smoking is regulated or prohibited.

Customs and Border Control: Applied to identify illegal tobacco smuggling.

Data Analytics and Research Organizations:

Data Analytics: Utilized by organizations for data-driven research on smoking behaviors and trends.

Environmental and Air Quality Monitoring:

Air Quality Assessment: Used to assess the impact of smoking in enclosed spaces and contribute to indoor air quality improvements.

The application of this technology in these various domains can contribute to reducing smoking-related health issues, improving public health, and fostering healthier behaviors among individuals and communities.

9. CONCLUSION

In conclusion, the project "Estimating Presence or Absence of Smoking through Bio Signals" represents a promising and innovative approach to addressing the challenge of identifying smoking habits with greater efficiency and accuracy. Through the application of machine learning algorithms and the analysis of bio signals, this project has the potential to revolutionize our ability to monitor and detect smoking behaviors. Here are some key takeaways:

Enhanced Accuracy: By harnessing the power of machine learning, this project can provide a more precise and reliable means of determining the presence or absence of smoking, reducing reliance on self-reporting and offering a comprehensive assessment of an individual's smoking habits.

Continuous Monitoring: The project's ability to analyze bio signals over time allows for long-term monitoring, making it invaluable for understanding the persistence and patterns of smoking behavior, beyond just detecting recent exposure.

Real-time Detection: Real-time detection capabilities open the door to immediate interventions and support for individuals attempting to quit smoking, potentially improving their chances of success.

Personalized Insights: Machine learning models can be fine-tuned to consider individual variations in bio signals, offering a more personalized approach to smoking detection, which can be especially useful given that smoking habits can vary widely among individuals.

Privacy Protection: The project can be designed with privacy in mind, ensuring that sensitive bio signal data is handled securely and ethically, alleviating concerns about data privacy.

10. FUTURE SCOPE

This innovative approach not only has immediate applications in smoking cessation programs and public health initiatives but also sets the stage for further developments in utilizing bio signals for health monitoring and behavior analysis.

In summary, the "Estimating Presence or Absence of Smoking through Bio Signals" project represents a significant step forward in our ability to combat smoking-related health issues by offering a more sophisticated, continuous, and accurate method of identifying smoking behaviors. Its potential impact extends to improving the well-being of individuals and contributing to broader public health goals.

11. BIBILOGRAPHY

https://en.wikipedia.org/wiki/Biosignal

https://en.wikipedia.org/wiki/Smoking_cessation

Books: "Machine Learning for Healthcare" by Pratap Dangeti

"Machine Learning in Medicine" by Tony L. Xing and Russell Greiner.

APPENDIX

Source Code

app.py

```
from flask import Flask,render_template,url_for,request,redirect,session

import pickle

import os

import re

from newsapi import NewsApiClient
```

```
app=Flask(__name__)
with open('Estimating The Presence or Absence of Smoking Through Bio-Signals.pkl', 'rb') as
f:
    model = pickle.load(f)
@app.route('/',methods=['POST','GET'])
def homepage():
    return render_template('index.html')
@app.route('/submit.html',methods=['POST','GET'])
def submit():
    return render_template('submit.html')
@app.route('/result',methods=['GET','POST'])
def Predict():
    if request.method =="POST":
        ID=request.form['ID']
        Gender=request.form['Gender']
        Age=request.form['Age']
        Height=request.form['Height(cm)']
        Weight=request.form['Weight(kg)']
        Eyesight_right=request.form['Eyesight (right)']
        Eyesight_left=request.form['Eyesight (left)']
        Hearing_right=request.form['Hearing (right)']
        Hearing_left=request.form['Hearing (left)']
        Waist=request.form['Waist(cm)']
        Hemoglobin=request.form['Hemoglobin']
        ALT=request.form['ALT']
        gtp=request.form['gtp']
        systolic=request.form['systolic']
        relaxation=request.form['relaxation']
        fasting_blood_sugar=request.form['fasting blood sugar']
        Cholesterol=request.form['Cholesterol']
        triglyceride=request.form[ 'triglyceride']
        urine_protein=request.form['urine protien']
```

```python
        HDL=request.form['HDL']
        LDL=request.form['LDL']
        oral=request.form['oral']
        dental_caries=request.form['dental caries']
        serum_creatinine=request.form['serum creatinine']
        AST=request.form['AST']
        tartar=request.form['tartar']
 Pred = [[float (ID), float (Gender), float (Age), float (Height), float (Weight), float
(Eyesight_right), float (Eyesight_left), float (Hearing_right),float(Hearing_left), float (Waist),
float (Hemoglobin), float (ALT), float (gtp), float(systolic), float (relaxation), float
(fasting_blood_sugar),float(Cholesterol),float(triglyceride),float(urine_protein), float (HDL),
float(LDL), float (oral), float(dental_caries), float(serum_creatinine), float
(AST),float(tartar)]]
    species=model.predict(Pred)[0]
    return render_template('/result.html',Predict=species)
if __name__=='__main__':
    app.run(debug=True)
```

index.html

```html
<!DOCTYPE html>
<html>
<head>
  <title>Estimation of Smoke</title>
  <link rel="stylesheet"  href="/static/style.css">
</head>
<body>
  <header>
    <h1>Estimation of Smoke</h1>
    <nav>
      <a href="/">Home</a>
      <a href="submit.html">Predict</a>
    </nav>
  </header>
  <main>
    <section id="hero">
```

```html
    <h2>Estimating the presence or absence of smoking through bio-signals</h2>
    <a href="#">Learn more</a>
   </section>
   <section id="about">
    <h3>About Us</h3>
    <p>We are a team of scientists and engineers developing innovative technologies to estimate the presence or absence of smoking through bio-signals. Our mission is to help people make informed decisions about their health and well-being.</p>
   </section>
   <section id="contact">
    <h3>Contact Us</h3>
    <p>We would love to hear from you! Please feel free to contact us with any questions or feedback.</p>
    <form action="mailto:info@estimationofsmoke.com" method="post">
     <input type="text" name="name" placeholder="Your Name">
     <input type="email" name="email" placeholder="Your Email">
     <input type="text" name="subject" placeholder="Subject">
     <textarea name="message" placeholder="Your Message"></textarea>
     <input type="submit" value="Send">
    </form>
   </section>
  </main>
  <footer>
  </footer>
</body> </html>
```

Result.html

```html
<!DOCTYPE html>
<html lang="en">
<head>
   <meta charset="UTF-8">
   <meta name="viewport" content="width=device-width, initial-scale=1.0">
   <title>SMOKING PREDICTION</title>
</head>
```

```
<body>
   <h1>THE Value Is: </h1>
   <h1>{{Predict}}</h1>
   <a href="/">Home</a>
</body>
</html>
```

Submit.html

```
<html>
   <head>
      <title>Estimation of Smoke</title>
      <link rel="stylesheet"  href="/static/style2.css">
      <style>
        .div1{
         background-color: rgb(135, 135, 143);
         float: left;
         width: 700px;
   }
        .div2{
         background-color: rgba(155, 12, 12, 0.451);
         float:right;
         width: 700px;
        }
      </style>
     </head>
     <body>
<form action="/result" method="post">
   <div class="div1">
   <fieldset>
     <legend>General Information</legend>
     <label for="ID">ID</label>
     <input name="ID" required>
     <label for="Gender">Gender</label>
```

```
  <input name="Gender" required>
  <label for="Age">Age</label>
  <input name="Age" required>
  <label for="Height">Height</label>
  <input name="Height(cm)">
  <label for="Weight">Weight</label>
  <input name="Weight(kg)">
  <label for="Eyesight_right">Eyesight_right</label>
  <input name="Eyesight (right)">
  <label for="Eyesigh_left">Eyesight_left</label>
  <input name="Eyesight (left)">
  <label for="Hearing_right">Hearing_right</label>
  <input name="Hearing (right)">
  <label for="Hearing_left">Hearing_left</label>
  <input name="Hearing (left)">
  <label for="Waist">Waist</label>
  <input name="Waist(cm)">
  <label for="Hemoglobin">Hemoglobin</label>
  <input name="Hemoglobin">
  <label for="ALT">ALT</label>
  <input name="ALT">
  <label for="gtp">gtp</label>
  <input name="gtp">
</fieldset></div>
<div class="div2">
<fieldset>
  <legend>Person Details</legend>
  <label for="systolic">systolic</label>
  <input name="systolic" required>
  <label for="relaxation">relaxation</label>
  <input name="relaxation" required>
  <label for="fasting blood sugar">fasting blood sugar</label>
```

```html
  <input name="fasting blood sugar" required>
  <label for="Cholesterol">Cholesterol</label>
  <input name="Cholesterol">
  <label for="triglyceride">triglyceride</label>
  <input name="triglyceride">
  <label for="urine protien">urine protien</label>
  <input name="urine protien">
  <label for="HDL">HDL</label>
  <input name="HDL">
  <label for="LDL">LDL</label>
  <input name="LDL">
  <label for="oral">oral</label>
  <input name="oral">
  <label for="dental caries">dental caries</label>
  <input name="dental caries">
  <label for="serum creatinine">serum creatinine</label>
  <input name="serum creatinine">
  <label for="AST">AST</label>
  <input name="AST">
  <label for="tartar">tartar</label>
  <input name="tartar">
</fieldset>
<input type="submit" value="Submit">
</div>  </form>  </html>
```